DOI: 10.17117/na.2016.05.03.075

<http://ucom.ru/doc/na.2016.05.03.075.pdf>

Поступила (Received): 02.05.2016

Заболотный А.С.**Информационная подсистема интеллектуального анализа данных на основе методов Data Mining****Zabolotny A.S.****Information subsystem data mining based on Data Mining methods**

В статье рассматривается возможность реализации информационной подсистемы интеллектуального анализа данных обучающихся на основе методов Data Mining, а также включены основные преимущества данных методов

Ключевые слова: информационная подсистема, интеллектуальный анализ данных

The article discusses the possibility of implementation of an information subsystem of data mining on the basis of participants during Data Mining methods, and also included the main advantages of these methods

Key words: information subsystem, data mining

Заболотный Александр Сергеевич

Бакалавр

Северо-Кавказский федеральный университет
г. Ставрополь, пр. Кулакова, 2**Zabolotny Alexandr Sergeevich**

Bachelor

North Caucasian federal university
Stavropol, Kulakov ave., 2

Научный руководитель:

к.физ-мат.н., доц. Крахоткина Е.В.

Быстротечное развитие информационных технологий, в частности, прогресс в методах сбора, хранения и обработки данных позволил многим организациям собирать огромные массивы данных, которые необходимо анализировать. Объемы этих данных настолько велики, что возможностей экспертов уже не хватает.

Сфера образования не стала исключением. Годами накапливалось множество информации о студентах, их успеваемости, преподавателях, создавались дистанционные курсы, образовательные форумы, системы тестирования студентов.

Существует множество задач, в которых методы статистики, машинного обучения и извлечения знаний (data mining) очень полезны как для учащихся, так и для преподавателя и для людей, ответственных за весь образовательный процесс в целом. Анализ данных позволяет лучше понять студентов, узнать какие предметы вызывают большие затруднения, как лучше построить курс, чтобы получить максимально высокие баллы, с какими тестами студенты справляются, какую форму занятий предпочитают, в какой области научных интересов преуспевают и т.д.

Data Mining переводится как "добыча" или "раскопка данных". Нередко рядом с Data Mining встречаются слова "обнаружение знаний в базах данных" (knowledge discovery in databases) и "интеллектуальный анализ данных". Их можно считать синонимами Data Mining. Возникновение всех указанных терминов связано с новым витком в развитии средств и методов обработки данных.

Интеллектуальный анализ данных образовательного процесса (EDM – Educational data mining) как раз занимается обработкой и анализом данных из образовательного процесса. Основные направления в данной области – использование анализа данных для поддержки интеллектуальных систем обучения (Intelligence Tutoring Systems), анализ образовательных процессов, визуализация данных образовательного процесса.

Целью данной статьи является исследование механизмов интеллектуального анализа данных (Data Mining) для решения задач извлечения знаний образовательного процесса и исследование возможностей Data Mining для реализации этих механизмов.

Для достижения поставленной цели необходимо решить следующие задачи:

- провести исследование основных методов EDM;
- рассмотреть задачи образовательного процесса, решение которых может быть найдено с использованием интеллектуального анализа данных;
- построить и реализовать алгоритмы Data Mining для решения задач образовательного процесса.

На первом этапе были рассмотрены основные методы интеллектуального анализа данных образовательного процесса. Краткое описание представлено в Таблице 1 [1].

Таблица 1. Основные методы интеллектуального анализа данных образовательного процесса

Метод	Описание метода
Предсказание (различают три типа предсказания – классификация, регрессия и оценка плотности)	Разработка модели, в которой вывод одиночного аспекта данных (предсказываемой переменной) осуществляется на основе комбинации других аспектов (переменных-предсказателей). Пример: предсказание, кто из студентов не справится с итоговым экзаменом
Кластеризация	Нахождение точек, которые естественным образом соединены вместе, разделение данных на кластеры. Пример: объединение студентов в группы, согласно их успеваемости по отдельным предметам
Анализ взаимоотношений	Изучение отношений между переменными в наборе из многих переменных
Исследование с помощью моделей	Изучение предварительно составленных моделей (разработанных с помощью EDM методов прогнозирования, кластеризации, инженерии знаний)
Преобразование данных к виду, понятному человеку	Преобразование сложных данных к виду, понятному для человека, для использования данных в человеческих суждениях

Первые три подхода – общие для data mining в различных областях, однако четвертый и пятый подходы относятся именно к анализу данных в образовательных доменах.

Одной из отличительных особенностей интеллектуального анализа данных образовательного процесса является то, что кроме методов data mining и машинного обучения здесь также применяются психометрические методы.

В процессе исследований было выявлено несколько задач, актуальных для университетов России, которые возможно решить при помощи интеллектуального анализа данных:

- выбор студентов для участия в международных олимпиадах и конкурсах;
- предоставление работодателям информации о наиболее подходящих для запрашиваемых должностей студентах-выпускниках;
- оценка вероятности трудоустройства студента по специальности, выявление степени влияния наличия сертификатов и побед в соревнованиях на трудоустройство.

Учитывая, что в большом количестве ВУЗов России информация о студентах хранится в СУБД Oracle и Microsoft SQL Server, целесообразным и наименее затратным представляется использование возможностей встроенных модулей этих СУБД.

Таким образом, анализ данных в образовательной сфере – новая и актуальная область науки, в которой существует множество нерешенных задач для анализа и исследования.

Список используемых источников:

1. Дюк В., Самойленко А. *Data Mining*. Спб.: Питер, 2001. 386 с.
2. Oracle Data Mining Concepts, Retrieved December 11, 2012.
URL: http://docs.oracle.com/cd/B19306_01/datamine.102/b14339/2data.htm
3. URL: http://www.i-teco.ru/Решения/Программные_решения/Отраслевые_решения/Решениядлясферыобразования
4. URL: <http://www.inftech.webservis.ru/it/database/datamining/ar2.html>

© 2016, Заболотный А.С.

Информационная подсистема интеллектуального анализа данных на основе методов Data Mining

© 2016, Zabolotny A.S.

Information subsystem data mining based on Data Mining methods